2D convolution model using (in)variant kernels for fast acoustic imaging

Ning Chu, Nicolas Gac, José Picheral, Ali Mohammad-Djafari

BeBeC-2014-05

2D CONVOLUTION MODEL USING (IN)VARIANT KERNELS FOR FAST ACOUSTIC IMAGING

Ning *CHU*[1], Nicolas *GAC*[1] and José *PICHERAL*[2], Ali *MOHAMMAD – DJAFARI*[1]
[1]Laboratoire des signaux et systèmes (L2S), CNRS-SUPELEC-PARIS SUD
[2]Département Signaux et Systèmes Electroniques (SSE), SUPELEC
91192 GIF-SUR-YVETTE, France

ABSTRACT

Acoustic imaging is an advanced technique for acoustic source localization and power reconstruction using limited measurements at microphone sensors. The acoustic imaging methods often involve in two aspects: one is to build up a forward model of acoustic power propagation which requires tremendous matrix multiplications due to large dimension of the power propagation matrix; the other is to solve an inverse problem which is usually ill-posed and time consuming. In this paper, our main contribution is to propose to use 2D convolution model for fast acoustic imaging. We find out that power propagation matrix seems to be a quasi-Symmetric Toeplitz Block Toeplitz (STBT) matrix in the far-field condition, so that the (in)variant convolution kernels (sizes and values) can be well derived from this STBT matrix. For method validation, we use simulated and real data from the wind tunnel S2A (France) experiment for acoustic imaging on vehicle surface.

1. INTRODUCTION

Acoustic imaging is an advanced technique for acoustic source localization and power reconstruction using limited measurements at microphone array. This technique can provide the insights into the performance, properties and mechanisms of acoustic sources. Nowadays, high-resolution acoustic imaging has been widely studied and applied in acoustic source reconstruction on the stationary, moving and rotating objects[12, 17]. Unfortunately, acoustic imaging often causes such an ill-conditioned inverse problem that solutions are not unique. Therefore, conventional methods cannot easily obtain a robust, efficient nor high resolution acoustic imaging.

According to the physical models, acoustic imaging methods could be generally classified into: time-reversal acoustic imaging [13], Near-field Acoustic Holography (NAH) [15] and inverse problems [19] etc. The latter one refers to using the measurements of forward model

for model parameter estimations. Moreover, inverse problems can be well solved by signal processing techniques and mathematical tools. In this paper, we mainly focus on the inverse problems which should account for the following two aspects:

- A forward model of acoustic propagation[12] including acoustic source model (monopole, extended or distributed), propagation paths (direct and indirect paths, reverberations), propagation types (near or far-field, full-wave or quasi-static analysis), and background noises (Gaussian white or colored or non-stationary distribution), .
- Its inverse problem[1] considering measured data, source spatial distribution, microphone array topology, and prior information on unknown parameters.

In general inverse problems, conventional Beamforming [4] method can give a fast and direct acoustic power imaging, but its spatial resolution is often low due to strong side lobe effects, especially for sources at low frequencies. In fact, Beamforming result can be interpreted as the source power image deteriorated by the 2D convolution caused by microphone array responses (convolution kernels). To deconvolve the Beamforming, the Deconvolution Approach for Mapping of Acoustic Source (DAMAS) method [2] has been proposed to effectively achieve high spatial resolutions. However, conventional DAMAS suffers from slow convergence because of different spatially-variant array responses. For fast convolution, extended DAMAS [6] assumes one spatially-invariant convolution kernel, but this assumption inevitably affects spatial resolutions. To overcome deconvolution drawbacks, Bayesian inference methods [1, 5, 14] have been a powerful methodology for solving ill-posed inverse problem. It can adaptively estimate both unknown random variables and unknown model parameters by applying the Bayes' rule in updating the probability law, in which, a posterior probability can be obtained from the likelihood and prior models. And the likelihood can be derived from forward model using measured data. The prior models can be assigned according to prior information on the unknowns, and the priors serve to promote useful regularizations on ill-posed inverse problems. Bayesian approach with Joint Maximum A Posterior (JMAP) criterion are usually used. However, Bayesian JMAP often causes tremendous computational burden due to non-quadratic or non-convex optimization. Above all, mentioned methods can be well performed for purpose use. But there is no universal methods fitting for all purposes.

In this paper, our motivation is to propose a fast acoustic imaging on the vehicle surface in wind tunnel tests, which can be practically used in automobile industry. The main contributions are: the power propagation matrix in the forward model is often a quasi-Symmetric Toeplitz Block Toeplitz (STBT) matrix in far-field condition; then this forward model can be approximated by using 2D convolution models with spatially-(in)variant convolution kernels (array responses), so that tremendous computation of matrix multiplication in original forward model can be greatly reduced.

This paper is organized as follows: Section 2 presents the classical forward model of acoustic propagation. Section 3 proposes 2D convolution models of acoustic power propagation and spatially-(in)variant convolution kernel selection. Section 4 and 5 validate the proposed 2D convolution approximation on simulations and real data respectively. Finally Section 6 concludes this paper.

2. Classic forward model

We assume that acoustic sources are uncorrelated monopoles [2, 6]; microphones are omni-directional with unitary gain; background noises at the microphones are Additive Gaussian White Noise (AGWN), independent and identically distributed (i.i.d); complex reverberation in the open wind tunnel could be neglected.

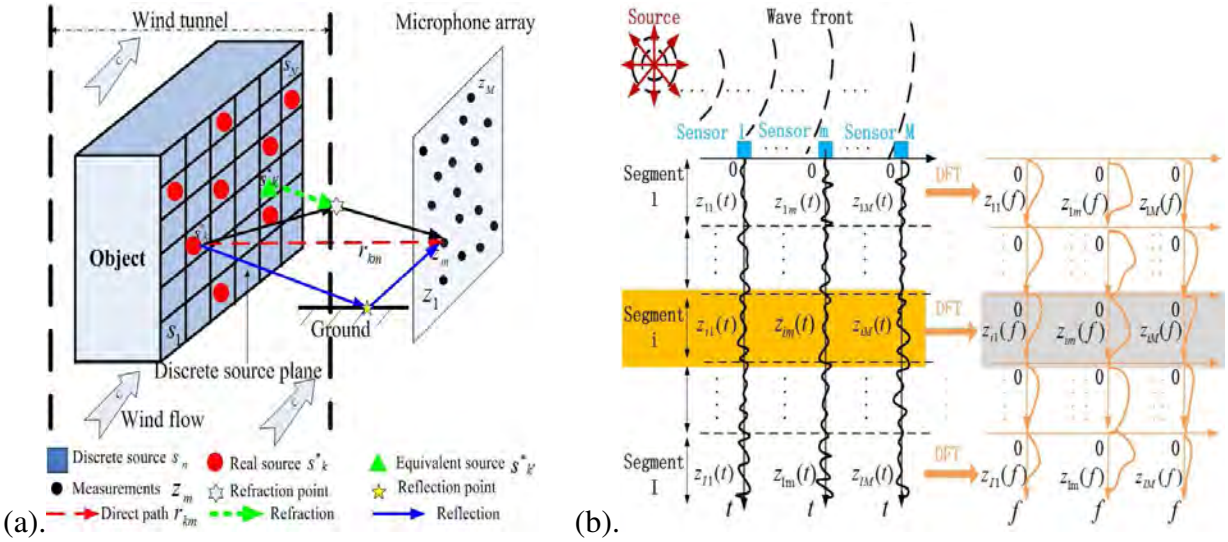


Figure 1: a. Illustration of the acoustic signal propagation in wind tunnel[5]. b. Illustration of the signal processing procedure in Eq.(1).

Figure 1(a) illustrates the acoustic signal propagation from the source plane to the microphone array in the wind tunnel, where microphones are installed outside the wind flow. On the source plane, we suppose K unknown original source signals $\mathbf{s}^* = [s_1^*, \cdots, s_K^*]^T$ at unknown positions $\mathbf{P}^* = [\mathbf{p}_1^*, \cdots, \mathbf{p}_K^*]^T$, where $\mathbf{p}_k^*$ denotes the 3D coordinates of kth original source signal s_k^*, notation $(\cdot)^*$ represents the original sources, and operator $(\cdot)^T$ denotes the transpose. On the microphone plane, we consider M microphones at known positions $\bar{\mathbf{P}} = [\bar{\mathbf{p}}_1, \cdots, \bar{\mathbf{p}}_M]^T$. The source plane is then equally discretized into N grids at known positions $\mathbf{P} = [\mathbf{p}_1, \cdots, \mathbf{p}_N]^T$. We assume that K original sources $\mathbf{s}^*$ sparsely distribute on these grids, satisfying $N > M > K$ and $\mathbf{P}$ including $\mathbf{P}^*$. We thus get N discrete source signals $\mathbf{s} = [s_1, \cdots, s_N]^T$ at known positions $\mathbf{P}$, satisfaying $s_n = s_k^*$, for $\mathbf{p}_n = \mathbf{p}_k^*$; $s_n = 0$ others. Since K<<N, $\mathbf{s}$ is full of zero, and it becomes a sparse signal with K-sparsity in the space domain. Therefore, to reconstruct $\mathbf{s}^*$ is to reconstruct K-sparsity signal $\mathbf{s}$. And $\mathbf{p}_k^*$ can be deprived from the discrete position $\mathbf{p}_n$, where s_n is non-zero.

2.1. Forward model of acoustic signal propagation

Signal processing procedure is illustrated in Fig.1(b). For the mth microphone with $m \in [1, \cdots, M]$, there are T samplings of acoustic signals in time domain. Then these T temporal samplings are divided into I blocks with L samplings in each block. We note $\mathbf{z}_{i,m}(t)$ as the received signal of the ith sampling block ($i \in [1, \cdots, I]$) at the mth microphone in the sampling

time $t \in [(i-1)L+1, \cdots, iL-1]$, and total sampling number is noted by $T = I \times L$. Since original source signals are usually of wide-band, we apply the Discrete Fourier Transform (DFT) in time domain to treat measured signals $\mathbf{z}_{i,m}(t)$ at each block so as to obtain L narrow frequency bins f_l (l$\in [1, \cdots, L]$). Let $\mathbf{z}_i(f_l) = [z_{i,1}(f_l), \cdots, z_{i,M}(f_l)]^T$ denote all measured signals in frequency domain. The signal processing is made independently for each frequency bin, thus in the following, we omit f_l for simplicity. Thus $\mathbf{z}_i$ can be modeled [2, 5, 20] as

$$\mathbf{z}_i = \mathbf{A}(\mathbf{P})\,\mathbf{s}_i + \mathbf{e}_i\,, \tag{1}$$

where $\mathbf{A}(\mathbf{P}) = [\mathbf{a}(\mathbf{p}_1)\cdots\mathbf{a}(\mathbf{p}_N)]$, $\mathbf{A}(\mathbf{P}) \in \mathbb{C}^{M\times N}$ consists of N steering vectors $\mathbf{a}(\mathbf{p}_n) = \left\{\frac{1}{r_{n,1}}\exp[-j(2\pi f_l \tau_{n,1})], \cdots, \frac{1}{r_{n,M}}\exp[-j(2\pi f_l \tau_{n,M})]\right\}^T$, with $r_{n,m}$ being the distance from source n to sensor m, $\tau_{n,m}$ propagation time during $r_{n,m}$. For $r_{n,m}$, we also consider the ground reflection and wind refraction in authors' paper [5]. For simplicity, $\mathbf{a}(\mathbf{p}_n)$ is short as $\mathbf{a}_n$ afterwards.

In summary, the forward model of signal propagation in Eq.(1) is a linear but underdetermined (M<N) system of equations for solving K-sparsity signal $\mathbf{s}$.

2.2. Forward model of acoustic power propagation

Based on Eq.(1), it is convenient to obtain the forward model of acoustic power propagation using Beamforming methods [2, 4, 5]:

$$\mathbf{y} = \mathbf{C}\,\mathbf{x} + \sigma_e^2\,\mathbf{1}_a\,, \tag{2}$$

where $\mathbf{y} = \{y_n\}_N^T$ denotes the Beamforming power vector; y_n can be interpreted as the estimated source power at grid n. And $\mathbf{y} = \tilde{\mathbf{A}}^\dagger\,\mathbb{E}[\mathbf{z}\mathbf{z}^\dagger]\,\tilde{\mathbf{A}}$ can be directly obtained from Eq.(1), where $\tilde{\mathbf{A}} = [\tilde{\mathbf{a}}(\mathbf{p}_1)\cdots\tilde{\mathbf{a}}(\mathbf{p}_N)]$, $\tilde{\mathbf{A}}(\mathbf{P}) \in \mathbb{C}^{M\times N}$ denotes the Beamforming steering matrix, and $\tilde{\mathbf{a}}(\mathbf{p}_n) = \frac{\mathbf{a}_n}{\|\mathbf{a}_n\|_2^2}$, operator $(\cdot)^\dagger$ denotes conjugate transpose, $\mathbb{E}[\cdot]$ denotes mathematical expectation. In practice, $\mathbb{E}[\mathbf{z}\mathbf{z}^\dagger] \approx \frac{1}{I}\sum_i^I \mathbf{z}_i\mathbf{z}_i^\dagger$ is approximated. $\boldsymbol{x} = diag\{\mathbb{E}[\boldsymbol{s}\boldsymbol{s}^H]\}$ denotes the unknown source power vector, and $diag\{\cdot\}$ denotes diagonal items; thus $\mathbf{x}$ is a signal as K-sparsity as $\mathbf{s}$. And σ_e^2 denotes the variance of i.i.d AGWN noises $\boldsymbol{e}$. Notation $\mathbf{1}_a = [\frac{1}{\|\mathbf{a}_1\|^2}, \cdots, \frac{1}{\|\mathbf{a}_N\|^2}]^T$ represents the noise attenuation for different grids. $\mathbf{C} = \{c_{i,j}\}_{N\times N}$ denotes the power propagation matrix, defined as:

$$\begin{aligned} c_{i,j} &= \frac{\|\mathbf{a}_i^H\mathbf{a}_j\|_2^2}{\|\mathbf{a}_i\|_2^2} \\ &= \left|\frac{1}{\sum_{m=1}^M \frac{1}{r_{im}^2}}\sum_{m=1}^M \frac{1}{r_{im}\,r_{jm}}\,e^{-j\frac{2\pi f_l}{c_0}(r_{jm}-r_{im})}\right|^2, \end{aligned} \tag{3}$$

where $\mathbf{a}_i$ is defined in Eq.(1); r_{im} denotes the propagation distance from ith discrete source (at the position $\mathbf{p}_i$ on the discrete source plane) to the mth microphone; f_l denotes the lth frequency bin; M is the total number of microphones. According to Eq.(3), it yields $0 \le c_{i,j} \le 1$ and $c_{i,i} = 1$. In fact, $c_{i,j}$ can represent the power contribution (%) of the microphone array from the jth source to the ith position on the source plane. So that $c_{i,j}$ can also be seen as the Point Spread Function (PSF) of the microphone array. This PSF is determined by two factors: the microphone array

topology and the distance from the source plane. In ideal case, $c_{i,j} = \delta_{i,j}$ becomes the Dirac function, and it derives $\mathbf{y} = \mathbf{x} + \sigma_e^2 \mathbf{1}_a$ from Eq.(2), which is easy to solve.

Compared with signal propagation model of Eq.(1), the power propagation model of Eq.(2) is a linear and determined system of equations for solving K-sparsity source powers $\mathbf{x}$.

2.3. Computational complexity in forward model

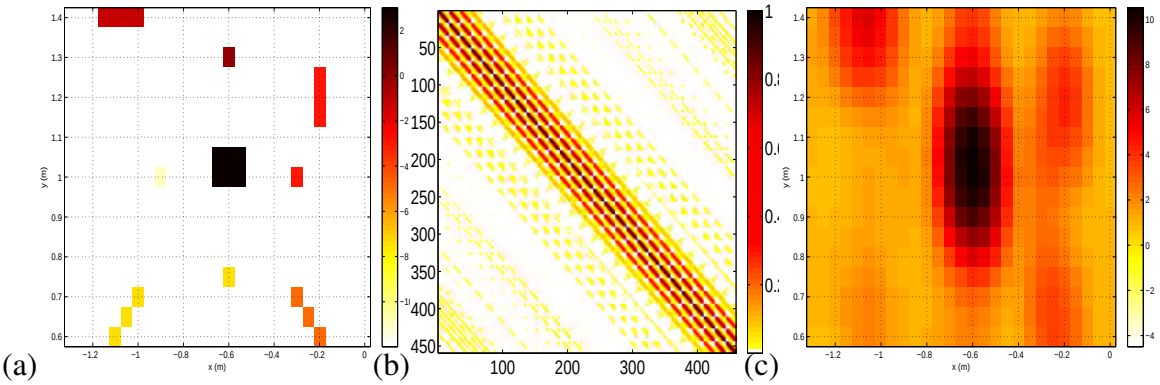


Figure 2: Simulation 1 on forward model of power propagation in Eq.(2), 23 monopole source, 14dB dynamic range among source powers, 15cm interval spaced, 5cm grid, 64 sensors, 4m averaged distance, 2500Hz working frequency, 0dB SNR, no reflection nor refraction: (a) Source power image $\mathbf{x}_0$ *(size:* 17×27*). (b) Power propagation matrix* $\mathbf{C}$ *(size:* 459×459*) (c) Measured Beamforming power image* $\mathbf{y}_0$ *(size:* 17×27*).*

In Fig.2, we show one example of Eq.(2). For the N-length vector $\mathbf{x}$ and $N \times N$ power propagation matrix $\mathbf{C}$, matrix multiplications $\mathbf{C}\mathbf{x}$ causes the computational complexity as heavy as $O(N^2)$. But $\mathbf{C}$ seems to be a quasi Symmetric Toeplitz Block Toeplitz matrix (STBT) [11]. In that case, $\mathbf{C}\mathbf{x} \approx \mathbf{h} * \mathbf{x}_0$ could be approximated to the 2D-convolution model, where $\mathbf{x}_0$ denotes the source power image, which is matrix form of vector $\mathbf{x}$; and $\mathbf{h}$ denotes the 2D invariant kernel, with the size of $N_h \times N_h$ ($N_h^2 < N$); operator $*$ denotes valid convolution: the output matrix of valid convolution consists of those overlap parts without zero-padded edges, so that the output matrix is the same size of input matrix. Owing to convolution approximation, the computational complexity can be significantly reduced from $O(N^2)$ into $O(N_h^2 N)$, even further $O(N log_2 N)$ using the Fast Fourier Transformation (FFT) [3]. In particular, if the 2D-convolution kernel can be separable into $\mathbf{h} = \mathbf{h}_1 * \mathbf{h}_2^T$, where $\mathbf{h}_1$ and $\mathbf{h}_2$ are vectors with N_h length. In that case, the computational complexity of matrix multiplications $\mathbf{C}\mathbf{x}$ can be greatly reduced into $O(2 N_h N)$. In brief, the computational complexity comparison is shown in Table 1.

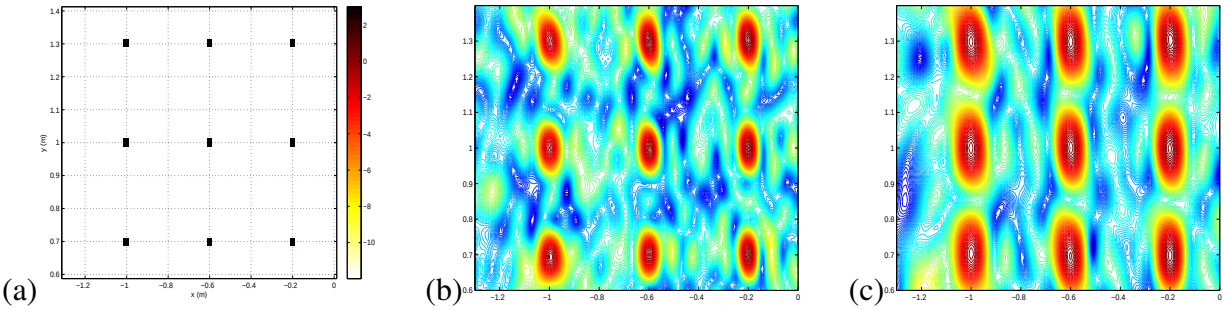
3. Proposed convolution models of acoustic power propagation

The power propagation model in Eq.(2) reveals that the source power reconstruction can be seen as the image deconvolution from the blurred Beamforming result. However, the Beamforming often involves in the convolutions with spatially variant kernels. This effect is shown in Fig. 3(a-b): same sources produce different shapes of PSFs on different positions, and the center

Table 1: The computational complexity comparison of different operators.

Operation	Expression	Complexity	Speed gain
Matrix multiplication	$\mathbf{C}\,\mathbf{x}$	$O(N^2)$	1
2D invariant convolution	$\mathbf{h} * \mathbf{x}_0$	$O(N_h^2 N)$	N/N_h^2
2D separable invariant convolution	$\mathbf{h}_1 * \mathbf{h}_2^T * \mathbf{x}_0$	$O(2N_h N)$	$N/2N_h$

PSF has the smallest size, while the ones on the corners have much larger sizes. But in Fig. 3(c), all the PSFs look like similar to each other in the far-field.



(a) (b) (c)

Figure 3: Simulation 2 for 9 monopole sources, 0dB dynamic range, 15cm interval, 2.5cm grid, 12000 pixels, 64 sensors, 2500Hz, no background noise, no reflection nor refraction: (a). Source power image (b). Spatially variant PSF in near-field condition (c). Approximated spatially invariant PSF in far-field condition

In order to derive the 2D-convolution model from Eq.(2), we firstly transform the vector $\mathbf{x}$ to matrix $\mathbf{x}_0$. We suppose that the source plane is discretized by $N = N_r \times N_c$ identical grids, where N_r and N_c denote row and column number respectively, provided $N_r \leq N_c$ for a rectangular plane. We then define the source power image $\mathbf{x}_0 = [x_{p,q}]_{N_r \times N_c}$ with $p \in [1, \cdots, N_r]$, $q \in [1, \cdots, N_c]$. Then $\mathbf{x}_0$ can be vectorized to $\mathbf{x} = [x_j]_N$ in column-first order as: $x_j = x_{p,q}\,, \quad j = p + (q-1)N_r$. So that for $\mathbf{y}_0 = [x_{p,q}]_{N_r \times N_c}$ and $\mathbf{y} = [y_i]_N$, we thus have $y_i = y_{p,q}\,, \quad i = p + (q-1)N_r$.

We then derive convolution kernels from the power propagation matrix $\mathbf{C}$ [7, 10]. In the far-field condition, $\mathbf{C}$ can be separated into a STBT matrix $\tilde{\mathbf{C}}$, and two diagonal matrices $\mathbf{D}_1, \mathbf{D}_2$ as: (See Appendix for details)

$$\mathbf{C} \approx \mathbf{D}_1 \tilde{\mathbf{C}} \mathbf{D}_2\,, \tag{4}$$

where $\mathbf{D}_1 = Diag\left[\bar{r}_i^2\right]$ and $\mathbf{D}_2 = Diag\left[\frac{1}{\bar{r}_j^2}\right]$ with $i, j \in [1, \cdots, N]$ denote the diagonal matrice; $\bar{r}_i = \frac{1}{M}\sum_{m=1}^{M} r_{i,m}$ denotes the averaged distance from the ith source to all microphone sensors; and $\tilde{c}_{i,j} \in \tilde{\mathbf{C}}$ is

$$\tilde{c}_{i,j} = \frac{1}{M^2}\left|\sum_{m=1}^{M} e^{j\frac{2\pi f_l}{c_0}(r_{i,m} - r_{j,m})}\right|^2, \tag{5}$$

Therefore, using Eq.(4) to replace $\mathbf{C}$ in the original forward model of Eq.(2), it yields

$$\tilde{\mathbf{y}} = \tilde{\mathbf{C}}\,\tilde{\mathbf{x}} + \boldsymbol{\epsilon}, \tag{6}$$

where $\tilde{\mathbf{y}} = \mathbf{D}_1^{-1}\,\mathbf{y}$ denotes the measured Beamforming power vector with attenuation $\mathbf{D}_1^{-1}$; and $\tilde{\mathbf{x}} = \mathbf{D}_2\,\mathbf{x}$ denotes the source power vector with attenuation $\mathbf{D}_2$; and $\boldsymbol{\epsilon} = \sigma_e^2\,\mathbf{D}_1^{-1}\,\mathbf{1}_a$ denotes the model errors. Since the averaged distance $\bar{r}_i$ can be easily calculated beforehand, we take $\tilde{\mathbf{x}}$ as $\mathbf{x}$ and $\tilde{\mathbf{y}}$ as $\mathbf{y}$ for symbol simplicity in the followings.

According to the STBT matrix $\tilde{\mathbf{C}}$, we can rewrite Eq.(6) by using the 2D convolution model as:

$$\mathbf{y} = \mathbf{H}\,\mathbf{x} + \boldsymbol{\epsilon}, \tag{7}$$

where matrix $\mathbf{H}$ denotes valid convolution matrix, satisfying:

$$[\mathbf{H}\,\mathbf{x}]_i = [\mathbf{h} * \mathbf{x}_0]_{p,q}, \quad i = p + (q-1)\,N_r, \tag{8}$$

where index $[\cdot]_i$ represents the ith item of a vector; index $[\cdot]_{p,q}$ represent the pth row, qth column item of a matrix; N_r denotes row size of the source plane. To express convolution kernel $\mathbf{h}$, we will discuss the spatially-variant and spatially-invariant two cases in this subsection.

3.1. 2D spatially-variant kernel

According to the STBT matrix $\tilde{\mathbf{C}}$ in Eq.(6), spatially-variant kernels in Eq.(8) can be derived as [8, 9, 18]:

$$\mathbf{h} = \sum_{p=1}^{N_c}\sum_{q=1}^{N_c} \mathcal{D}^{p,q}\,\mathbf{h}^{p,q}, \tag{9}$$

where $\mathcal{D}^{p,q}$ denotes the piecewise constant interpolation function [18], which is non-negative diagonal matrix satisfying $\sum_{p=1}^{N_c}\sum_{q=1}^{N_c}\mathcal{D}^{p,q} = \mathbf{I}$ (identity matrix), and the lth diagonal item is 1 if the lth PSF is in the region of (p,q). We call $\mathbf{h}^{p,q}$ the spatially-variant kernel, since $\mathbf{h}^{p,q}$ varies along with the convolution output $y_i \in \mathbf{y}$, $i = p + (q-1)\,N_r$ in Eq.(7).

According to the expression of $\tilde{\mathbf{C}}$ in Eq.(5), each item $h^{p,q}(k,l) \in \mathbf{h}^{p,q}$ in Eq.(9) is obtained as:

$$\begin{cases} h^{p,q}(k,l) = \tilde{c}_{i,j}, \\ i = p + (q-1)\,N_r, \quad j = i + (\lfloor \frac{N_r^h+1}{2} \rfloor - k)\,N_r + \lfloor \frac{N_c^h+1}{2} \rfloor - l \end{cases}, \tag{10}$$

where $N_r^h \times N_c^h$ denotes the kernel size; $k \in [1,\cdots,N_r^h]$, $l \in [1,\cdots,N_c^h]$; operator $\lfloor\cdot\rfloor$ denotes integer part.

In brief, $h^{p,q}(k,l)$ is derived from $\tilde{c}_{i,j}$ in three steps: Firstly, $\mathbf{h}^{p,q}$ comes from the specific $\tilde{c}_{i,j}$ which are on the same row of $i = p + (q-1)\,N_r$ in matrix $\mathbf{C}$; Then, the item $h^{p,q}(k,l)$ is derived by specific $c_{i,j}$ on the column j which is determined by the known index i,k,l as shown in Eq.(10); Finally, $\mathbf{h}^{p,q}$ should be flipped up-down, left-right according to the definition of 2D valid convolution.

3.2. 2D invariant kernel

Owing to the STBT matrix $\tilde{\mathbf{C}}$, its middle row ($i = \lfloor \frac{N+1}{2} \rfloor$) contains most of the useful items of other rows in $\tilde{\mathbf{C}}$. According to variant kernels in Eq.(10), we can derive an invariant convolution kernel $\mathbf{h} = [h_{k,l}]$ with $k, l \in [1, \cdots, N_r]$ from the middle row of $\tilde{\mathbf{C}}$ as:

$$\begin{cases} h_{k,l} = \tilde{c}_{i,j}, \\ i = \lfloor \frac{N+1}{2} \rfloor, \quad j = i + (\lfloor \frac{N_r+1}{2} \rfloor - k) N_r + \lfloor \frac{N_r+1}{2} \rfloor - l \end{cases}, \tag{11}$$

where $\mathbf{h}$ is can be a $N_r \times N_r$ square matrix, since the STBT matrix $\tilde{\mathbf{C}}$ consists of $N_r \times N_r$ square subblocks. Compared with the 'variant' kernel in Eq.(10), the 'invariant' kernel in Eq.(11) does not change along with convolution output y_i, $i = p + (q-1)N_r$, but remains the same $i = \lfloor \frac{N+1}{2} \rfloor$.

4. Simulations for 2D convolution model validation

On simulations, we will show the five following aspects: Approximation errors between STBT matrix $\tilde{\mathbf{C}}$ and power propagation matrix $\mathbf{C}$; Convolution approximated errors for variant, invariant and separable kernels, as well as different kernel sizes and forms (square or rectangular); Convolution computational time for different kernels; Acoustic imaging results based on 2D invariant convolution model.

The simulation configurations are shown in above figure: there are $M = 64$ non-uniform microphones locating on the vertical plane. $d = 2$m is the averaged size of microphone array. $D = 4.50$m is the distance between the microphone plane and source plane. $c_0 \approx 340$m/s is the acoustic speed in the common air. $T = 10000$ is the total number of samplings. For the simulated sources in Fig.2(a), there are simulated 4 monopoles and 5 complex sources, spaced at least 20cm from each other. Original source powers $\mathbf{x}^*$ are within [0.08,2] ([-10.3,3.7]dB) and 14dB dynamic range. And power image size is of $N_c = 27$ and $N_r = 17$ as shown in Fig.2(c). The i.i.d AWGN noise power is set $\sigma_e^2 = 0.86$ (-0.7dB), thus the averaged SNR is 0dB.

4.1. Approximation errors of STBT matrix

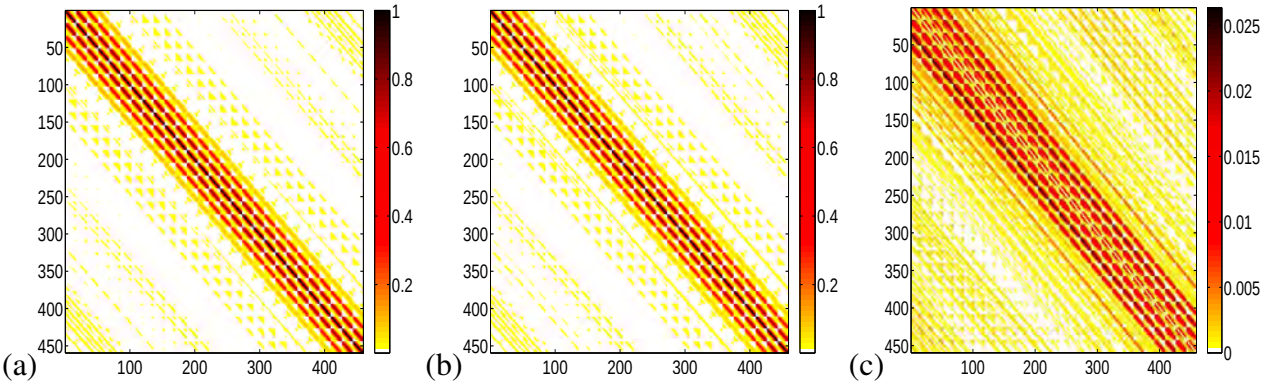


Figure 4: Power propagation matrix and its STBT approximation at 2500Hz: (a) $\mathbf{C} = [c_{i,j}]$ *(b)* $\tilde{\mathbf{C}} = [\tilde{c}_{i,j}]$ *(c) Approximation error matrix* $|c_{i,j} - \tilde{c}_{i,j}|$

In Fig.4, the matrix structures of $\mathbf{C} = [c_{i,j}]$ in Eq.(3) and $\tilde{\mathbf{C}} = [\tilde{c}_{i,j}]$ in Eq.(5) are quite similar to each other. The biggest relative error of STBT approximation is less than 2.5%. Therefore, the power propagation matrix $\mathbf{C}$ can be effectively approximated by the STBT matrix $\tilde{\mathbf{C}}$.

4.2. Convolution approximated errors for different kernels

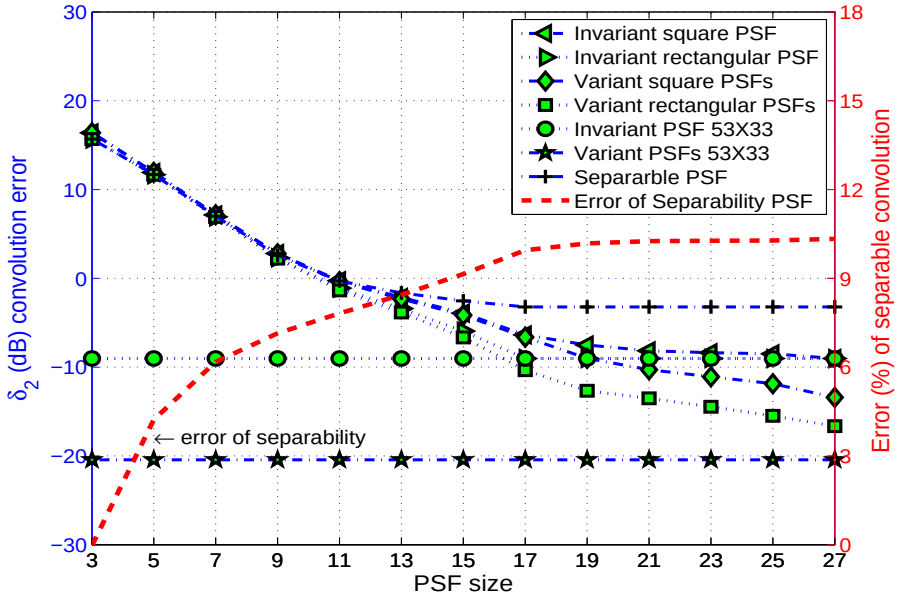


Figure 5: Convolution performance comparisons among variant, invariant and separable convolution kernels at 2500Hz.

We define the convolution approximated errors as $\delta_y = \frac{\|\mathbf{y}-\hat{\mathbf{y}}\|_2^2}{\|\mathbf{y}\|_2^2} \times 100\%$, where $\mathbf{y}$ refers to the Beamforming result using measured signals in Eq.(1) as shown in Fig.2(c); $\hat{\mathbf{y}}$ refers to the convolution results respectively using variant kernels in Eq.(10) and invariant kernel in Eq.(11).

In Fig.5, we show convolution approximated errors δ_y versus various kernel sizes. We examine 7 types of kernels with different forms. Firstly, both the variant and invariant kernel with the largest size of 53×33 obtain the very small convolution errors, which validates our proposed (in)variant convolution models in Eq.(10) and Eq.(11). Secondly, the larger kernel size is, the smaller δ_y becomes, and both the square and rectangular kernels obtain similar δ_y for each case, so that we can choose the square kernel for simplicity. Thirdly, the invariant kernel obtains as small δ_y as those of variant kernels, so that we can use invariant model to effectively approximate power propagation model. Fourthly, when kernel size N_h approaches source power image $N_r = 17$, all δ_y of 7 kernels becomes small and remains stable.

4.3. Acoustic imaging via 2D invariant convolution model

In the 2D convolution model of Eq.(6), this linear system of equation with STBT matrix $\tilde{\mathbf{C}}$ can be generally solved by multichannel Levinson algorithm [8, 9]. But in order to obtain robust source reconstruction in the strong background noises and approximation errors,

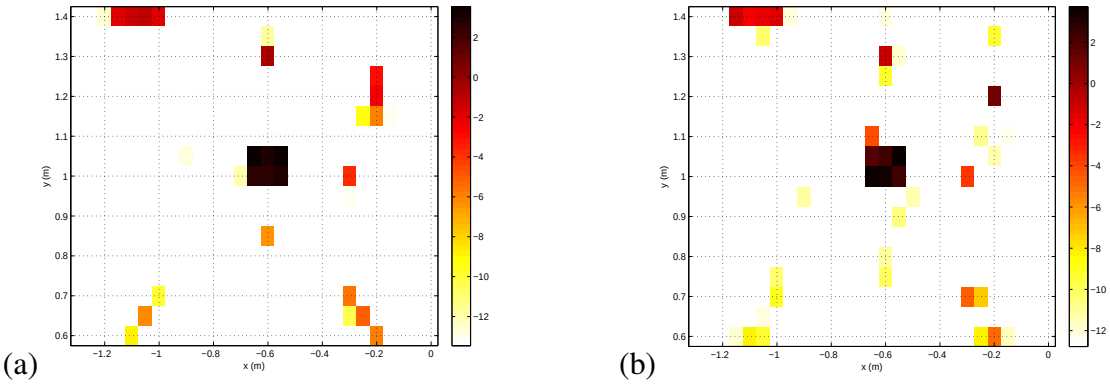


Figure 6: Simulations at 2500Hz, 0dB SNR, 15dB display: (a) Bayesian JMAP method via conventional forward model (b) Bayesian JMAP method via invariant convolution model

we apply the Bayesian JMAP method proposed in authors' paper [5] which aims to solve $\mathcal{J}(\mathbf{x},\boldsymbol{\theta}) = \|\mathbf{y}-\mathbf{H}\,\mathbf{x}-\boldsymbol{\epsilon}\|^2 + \mathcal{D}(\mathbf{x},\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ denotes prior model parameters; $\mathcal{D}(\cdot)$ denotes the regularization using the prior models on the unknowns.

In Fig.2, the Beamforming merely gives some strong source powers. In Fig.6, Bayesian JMAP method via classical forward model well detects all source powers except for the weakest monopole source. The Bayesian JMAP method via convolution model can quickly reconstruct most of the sources, but the recovered source patterns are affected by measured errors.

5. Wind tunnel experiments

Figure 7: Acoustic imaging on the vehicle surface in Wind tunnel S2A.

Figure 7 shows the configurations of the wind tunnel S2A [16], object vehicle, Non-uniform array and wind refraction. We suppose that all acoustic sources locate on the same plane. This assumption is almost satisfied, because the curvature of the car side is relatively small compared to the distance D=4.5m between the car and array plane. Since the scanning step is set by $\Delta p = 5cm$, the source plane of car side is of 1.5× 5 m^2 (31×101 pixels). On the real data, there

are T=524288 samplings with the sampling frequency f_s=2.56×10^4 Hz. We separate these samplings into I=204 blocks with L=2560 samplings in each bloc. The working frequency is 2500Hz which is sensitive to human being. The image results are shown by normalized dB images with 10dB span. For the actual propagation distance $r_{n,m}$ in Eq.(1), we apply equivalent source to make refraction correction, and the mirror source signal to correct the ground reflection as discussed in author's paper[5].

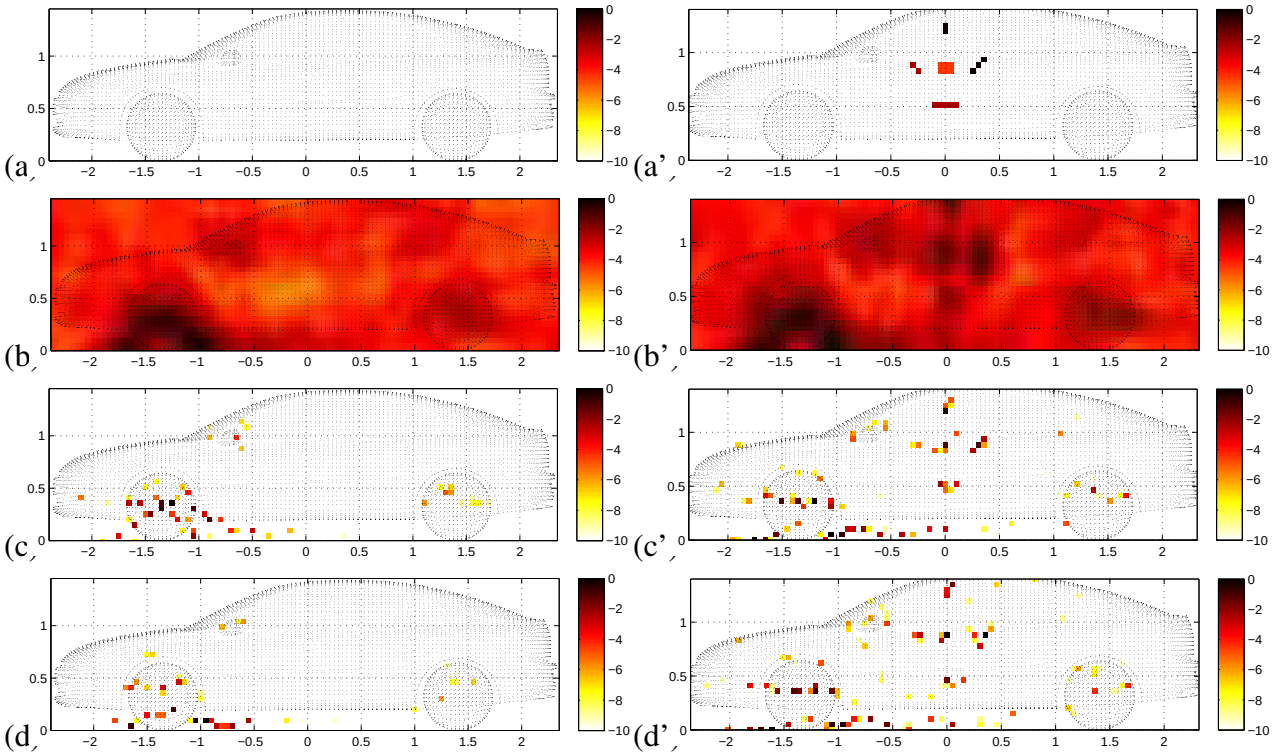


Figure 8: Left: real data at 2500Hz (a) vehicle surface (b) Beamforming (c) Bayesian JMAP via classical forward model (d) JMAP via invariant convolution model. Right: hybrid data (a') 5 simulated complex sources (b')-(d') corresponding methods.

Figure.8(a-d) illustrate the estimated power images of mentioned methods at 2500Hz. Due to the high sidelobe effect, Beamforming just gives a coarse image of strong sources. The Bayesian JMAP method via sparse prior not only manages to distinguish the strong sources around the two wheels, rearview mirror and side window, but also successfully reconstructs the week ones on the front cover and light. The Bayesian JMAP method via proposed 2D invariant convolution model can achieve the source reconstruction as good as the JMAP via conventional forward model. Figure.8(a'-d') use the hybrid data which simulated source are added to the real data. Bayesian via convolution model can further effectively detect both the simulated and original source powers in the real data.

In Table 2, the computation speed is greatly improved by 2D invariant convolution model in Eq.(11) compared to conventional model in Eq.(2)

Table 2: Computational cost for treating real data of whole car: image 31×101 pixels, at 2500Hz, based on CPU: 3.33Hz. 'JMAP+Conv' is short for Bayesian JMAP method via 2D invariant convolution model

Methods	CB	JMAP	JMAP+Conv
Time (s)	1	1012	180

6. Conclusions and perspectives

In this paper, we propose a 2D convolution model in Eq.(7) to approximate the forward model of source power propagation in Eq.(2), so that proposed Bayesian JMAP method is more quickly carried out. We firstly discuss the 2D-convolution model using a variant kernel in Eq.(10) and invariant kernel in Eq.(11) respectively. Convolution kernels (size and item values) are derived from the Symmetric Toeplitz Block Toeplitz (STBT) structure of power propagation matrix.

On simulations, the main conclusions are:

- There are relatively small approximation errors between STBT matrix $\tilde{\mathbf{C}}$ and power propagation matrix $\mathbf{C}$;
- 2D invariant convolution model successfully approximates the power propagation model;
- 2D invariant kernel whose size is just the half of the source power image efficiently performs the 2D convolution model.
- Bayesian JMAP method via 2D invariant convolution model obtains an acceptable imaging result compared to the conventional power propagation model;

On real data and hybrid data, we demonstrate that using 2D invariant convolution model can greatly accelerate the Bayesian JMAP method and contribute a rapid implementation for industry application.

However, there are at least three aspects to be further improved:

- The acoustic image quality using the 2D invariant or separable convolution model should be carefully refined, and it is necessary to balance between source reconstruction and computational cost, especially for real data treatment of wind tunnel experiments.
- To improve the source estimation results, the model error $\boldsymbol{\epsilon}$ in the proposed convolution model might not be always Gaussian white noise distribution, but probably the spatially non-stationary Gaussian distribution on the different parts of source plane or on the different microphones.
- For GPU implementation on the large scale of real data in tunnel experiments, it is quite worthy of optimizing advanced deconvolution algorithms (such as the Bayesian inference) via separable convolution model, so that the peak-power computation of GPU can be well utilized, and the drawbacks of limited local on-chip memory could be avoided to some extend.

A. STBT Matrix Approximation

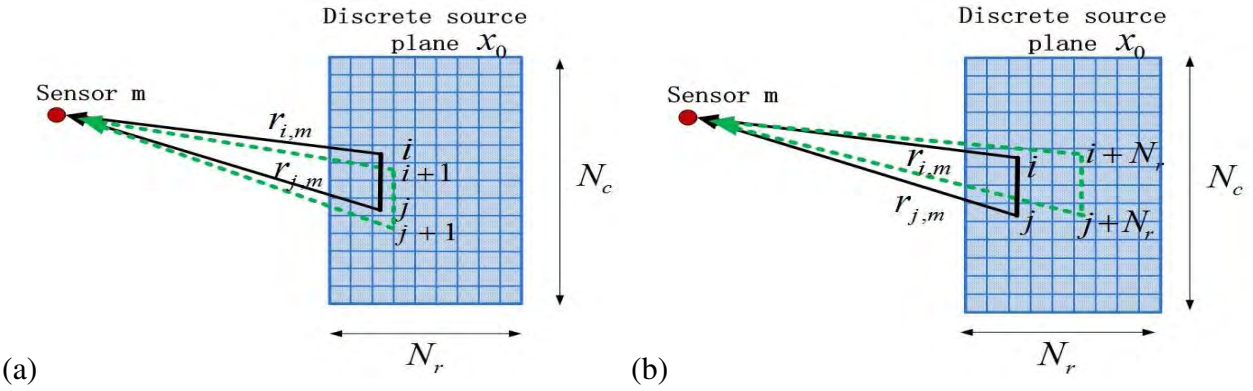


Figure 9: Assumptions for STBT matrix $\hat{\mathbf{C}}$ approximation (a) Approximation for Toeplitz in a subblock (b) Approximation for block Toeplitz block

For the ith source $i \in [1,\cdots,N]$, we suppose that there exists an averaged distance $\bar{r}_i = \frac{1}{M}\sum_{m=1}^{M} r_{i,m}$ from sources to the sensor plane, satisfying $\bar{r}_i / r_{i,m} \approx 1$ for any sensor $m \in [1,\cdots,M]$. According to the above assumption, each item $c_{i,j} \in \mathbf{C}$ in Eq.(3) can be approximated by:

$$\begin{aligned}\hat{c}_{i,j} &\approx \left| \frac{1}{\sum_{m=1}^{M} \frac{1}{r_{im}^2}} \sum_{m=1}^{M} \frac{1}{r_{im} r_{jm}} e^{j\frac{2\pi f_l}{c_0}(r_{im}-r_{jm})} \right|^2 \\ &= \bar{r}_i^2 \frac{1}{M^2} \left| \sum_{m=1}^{M} e^{j\frac{2\pi f_l}{c_0}(r_{i,m}-r_{j,m})} \right|^2 \frac{1}{\bar{r}_j^2}, \end{aligned} \tag{12}$$

According to Eq.(5), we have

$$\tilde{c}_{i,j} = \frac{1}{M^2} \left| \sum_{m=1}^{M} e^{j\frac{2\pi f_l}{c_0}(r_{i,m}-r_{j,m})} \right|^2, \tag{13}$$

where r_{im} denotes the propagation distance from ith discrete source (at the position $\mathbf{p}_i$ on the discrete source plane) to the mth sensor; f_l denotes the lth frequency bin; M is the total number of sensors; $i, j \in [1,\cdots,N]$; and c_0 is the acoustic propagation speed.

According to Eq.(13), we get $\tilde{c}_{i,j} = \tilde{c}_{j,i}$, $\tilde{\mathbf{C}} = \tilde{\mathbf{C}}^T$. Therefore, $\tilde{\mathbf{C}}$ is a symmetric matrix. And $\tilde{\mathbf{C}}$ can be expressed by subblock matrices $\tilde{\mathbf{C}}_{q,l}$ as follows:

$$\begin{cases} \tilde{\mathbf{C}} = [\tilde{\mathbf{C}}_{q,l}]_{N_c \times N_c} \quad , \tilde{\mathbf{C}}_{q,l} = [\tilde{c}_{p,k}^{(q,l)}]_{N_r \times N_r}, \quad \tilde{c}_{p,k}^{(q,l)} = \tilde{c}_{i,j} \in \tilde{\mathbf{C}} \\ i = p + (q-1)N_r,\ j = k + (l-1)N_r \end{cases} \tag{14}$$

where $\tilde{\mathbf{C}}_{q,l}$ with $q,l \in [1,\cdots,N_c]$ denotes the subblock matrix at qth-row and lth-column block of $\tilde{\mathbf{C}}$; and $\tilde{\mathbf{C}}$ has the number of $N_c \times N_c$ subbolcks. $c_{p,k}^{(q,l)}$ with $p,k \in [1,\cdots,N_r]$ denotes the pth-row and kth-column item of $\tilde{\mathbf{C}}_{q,l}$, and $\tilde{\mathbf{C}}_{q,l}$ has the size of $N_r \times N_r$.

We then suppose that

$$|r_{i,m} - r_{j,m}| \approx |r_{i+1,m} - r_{j+1,m}|, \quad \lfloor \frac{i}{N_r} \rfloor = \lfloor \frac{j}{N_r} \rfloor \quad ,(q = l) \tag{15}$$

where $\lfloor \cdot \rfloor$ denotes the integer part, which reflects that the ith and jth, $i+1$th and $j+1$th discrete sources are on the same column on the source power image.

Based on Eq.(13), it yields $\tilde{c}_{i,j} = \tilde{c}_{i+1,j+1}$ for i and j belong to one subblock. Since index i and j are periodically changing, we then have $\tilde{c}_{i,j} = \tilde{c}_{i+N_r,j+N_r}$ in two subblock.

According to Eq.(14) and (15), for any $\tilde{c}^{(q,l)}_{p,k}, \tilde{c}^{(q,l)}_{p+1,k+1} \in \tilde{\mathbf{C}}_{q,l}$, we have $\tilde{c}^{(q,l)}_{p+1,k+1} = \tilde{c}_{i+1,j+1} = \tilde{c}_{i,j} = \tilde{c}^{(q,l)}_{p,k}$. Therefore, subblock $\tilde{\mathbf{C}}_{q,l}$ is a Toeplitz matrix.

For any $\tilde{c}^{(q,l)}_{p,k} \in \tilde{\mathbf{C}}_{q,l}$ and $\tilde{c}^{(q+1,l+1)}_{p,k} \in \tilde{\mathbf{C}}_{q+1,l+1}$, we get $\tilde{c}^{(q+1,l+1)}_{k,l} = \tilde{c}_{i+N_r,j+N_r} = \tilde{c}_{i,j} = \tilde{c}^{(q,l)}_{p,k}$, and $\tilde{\mathbf{C}}_{q,l} = \tilde{\mathbf{C}}_{q+1,l+1}$. Therefore, $\tilde{\mathbf{C}}$ is a block Toeplitz matrix.

Above all, $\tilde{\mathbf{C}}$ is proved ton be a STBT matrix..

References

[1] J. Antoni. "A Bayesian approach to sound source reconstruction: optimal basis, regularization, and focusing." *The Journal of the Acoustical Society of America*, 131, 2873–2890, 2012.

[2] T. Brooks and W. Humphreys. "A Deconvolution Approach for the Mapping of Acoustic Sources (DAMAS) determined from phased microphone arrays." *Journal of Sound and Vibration*, 294(4-5), 856–879, 2006. ISSN 0022-460X.

[3] R. H. Chan, J. G. Nagy, and R. J. Plemmons. "FFT-based preconditioners for Toeplitz-block least squares problems." *SIAM journal on numerical analysis*, 30(6), 1740–1768, 1993.

[4] J. Chen, K. Yao, and R. Hudson. "Source localization and beamforming." *Signal Processing Magazine, IEEE*, 19(2), 30–39, 2002.

[5] N. Chu, A. Mohammad-Djafari, and J. Picheral. "Robust Bayesian super-resolution approach via sparsity enforcing a priori for near-field aeroacoustic source imaging." *Journal of Sound and Vibration*, 332(18), 4369–4389, 2013. ISSN 0022-460X.

[6] R. Dougherty. "Extensions of DAMAS and Benefits and Limitations of Deconvolution in Beamforming." In *11th AIAA/CEAS Aeroacoustics Conference*, pages 1–13. Monterey, CA, USA, 23-25 May, 2005. ISSN 0146-3705.

[7] R. M. Gray. *Toeplitz and circulant matrices: A review*. Now Pub, 2006.

[8] P. C. Hansen. "Deconvolution and regularization with Toeplitz matrices." *Numerical Algorithms*, 29(4), 323–378, 2002.

[9] J. Haupt, W. U. Bajwa, G. Raz, and R. Nowak. "Toeplitz compressed sensing matrices with applications to sparse channel estimation." *Information Theory, IEEE Transactions on*, 56(11), 5862–5875, 2010.

[10] M. Kac. "Toeplitz matrices, translation kernels and a related problem in probability theory." *Duke Mathematical Journal*, 21(3), 501–509, 1954.

[11] T. Kailath and J. Chun. "Generalized displacement structure for block-toeplitz, toeplitz-block, and toeplitz-derived matrices." *SIAM Journal on Matrix Analysis and Applications*, 15(1), 114–128, 1994.

[12] J. Lanslots, F. Deblauwe, and K. Janssens. "Selecting Sound Source Localization Techniques for Industrial Applications." *Sound and Vibration*, 44(6), 6–10, 2010. ISSN 1541-0161.

[13] S. Lehman and A. Devaney. "Transmission mode time-reversal super-resolution imaging." *The Journal of the Acoustical Society of America*, 113(5), 2742–2753, 2003.

[14] A. Massa and G. Oliveri. "Bayesian compressive sampling for pattern synthesis with maximally sparse non-uniform linear arrays." *IEEE Transactions on Antennas and Propagation*, 59(10), 467–681, 2011.

[15] J. D. Maynard, E. G. Williams, and Y. Lee. "Nearfield acoustic holography: I. Theory of generalized holography and the development of NAH." 78(4), 1395–1413, 1985. ISSN 00014966.

[16] A. Menoret, N. Gorilliot, and J.-L. Adam. "Acoustic imaging in wind tunnel S2A." In *10th Acoustics conference (ACOUSTICS2010)*. Lyon, France, 2010.

[17] A. B. Nagy. "Aeroacoustics research in Europe: The CEAS-ASC report on 2010 highlights." *Journal of Sound and Vibration*, 330(21), 4955–4980, 2011.

[18] J. G. Nagy and D. P. O'leary. "Fast iterative image restoration with a spatially varying psf." In *Optical Science, Engineering and Instrumentation'97*, pages 388–399. International Society for Optics and Photonics, 1997.

[19] A. Tarantola. *Inverse problem theory and methods for model parameter estimation*. Society for Industrial Mathematics, 2005.

[20] Y. Wang, J. Li, P. Stoica, M. Sheplak, and T. Nishida. "Wideband RELAX and wideband CLEAN for aeroacoustic imaging." *Journal of Acoustical Society of America*, 115(2), 757–767, 2004. ISSN 00014966.